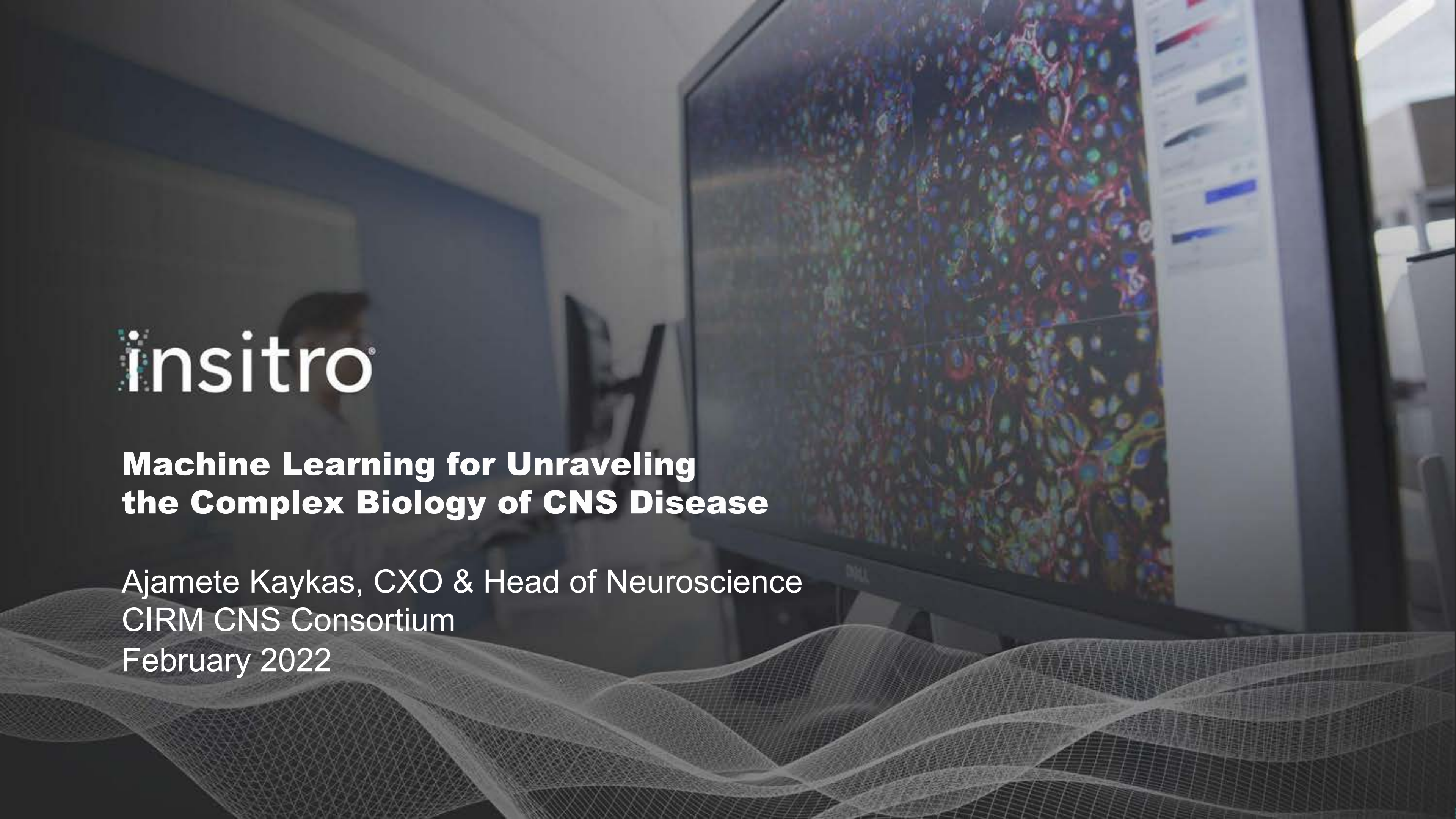


insitro

Machine Learning for Unraveling the Complex Biology of CNS Disease

Ajamete Kaykas, CXO & Head of Neuroscience
CIRM CNS Consortium
February 2022

We are insitro

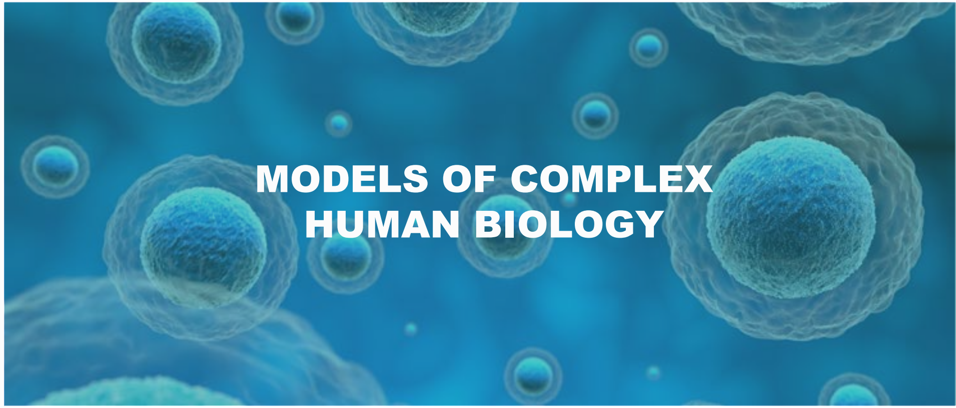
insitro

a seamless integration of ***in silico*** *and* ***in vitro***

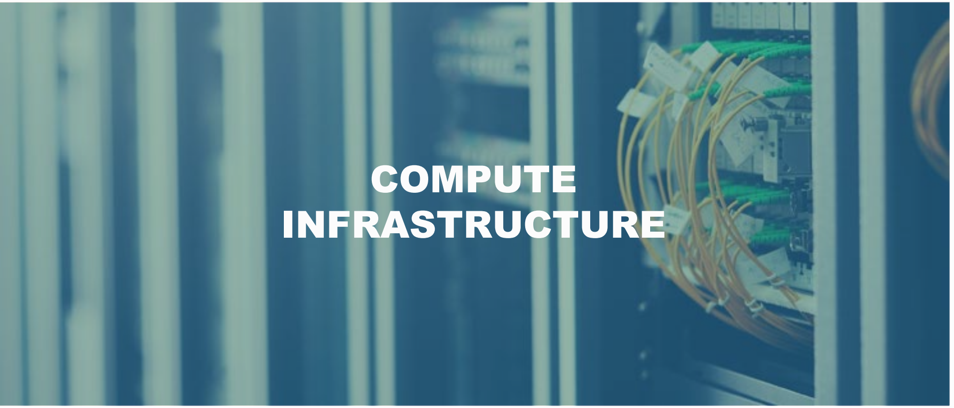
We are a drug discovery and development company enabled by a powerful convergence of machine learning and data generation at scale to bring medicines to the patients who need them

Multiple Forces Enable Our Transformative Approach

PATIENT-LEVEL DATA



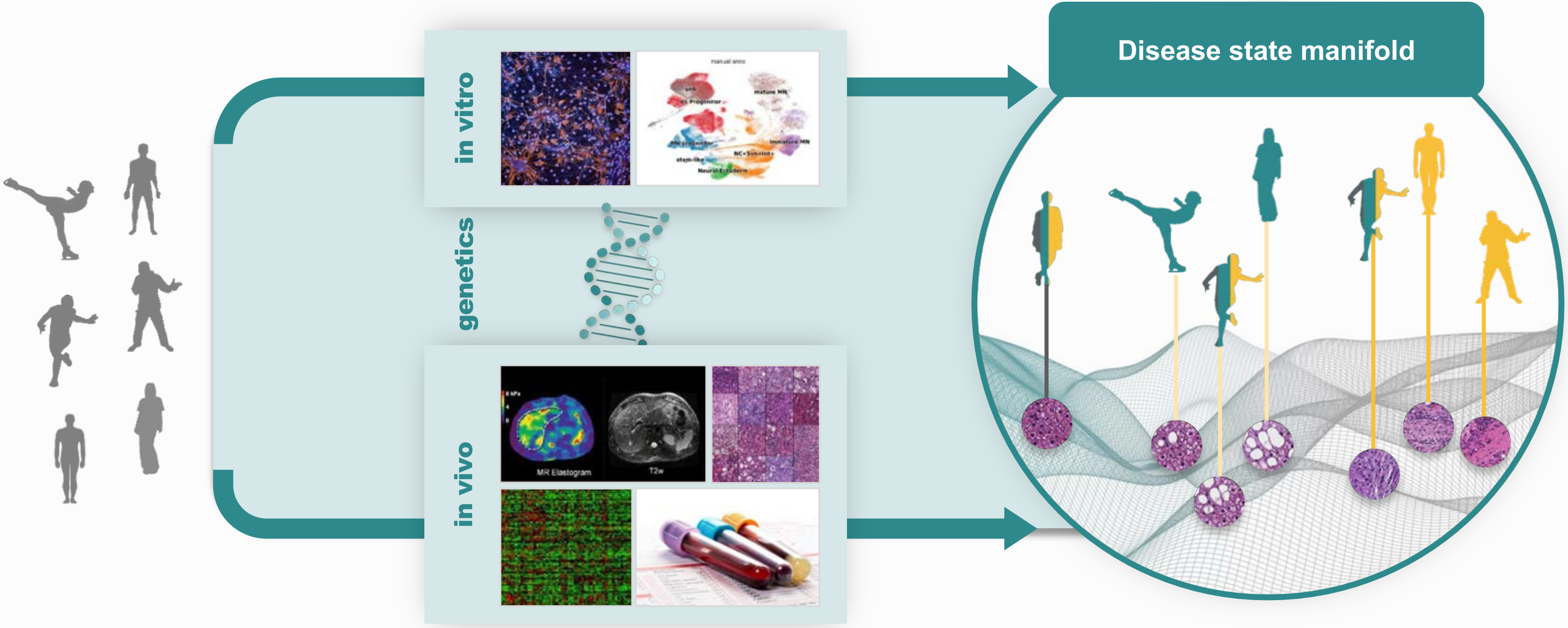
MODELS OF COMPLEX HUMAN BIOLOGY



COMPUTE INFRASTRUCTURE

MACHINE LEARNING

The insitro Platform Bridges Genetics and Clinical Outcomes with High-content Data



insitro Target and Drug Discovery Efforts

	Translational ML	Predictive Cell Models	Molecular Design
DRUG DISCOVERY	Translational ML (TML) and ML-enabled statistical genetics for patient state	Predictive cell-based disease models in the insitro Human Platform (ISH) for cellular state	ML-enabled therapeutics design with DNA-Encoded Libraries (DEL) and cellular screening
PREDICTIVE INSIGHTS	Target Patient	Target Patient	Molecule

insitro’s Efforts in Neuroscience

NEUROSCIENCE

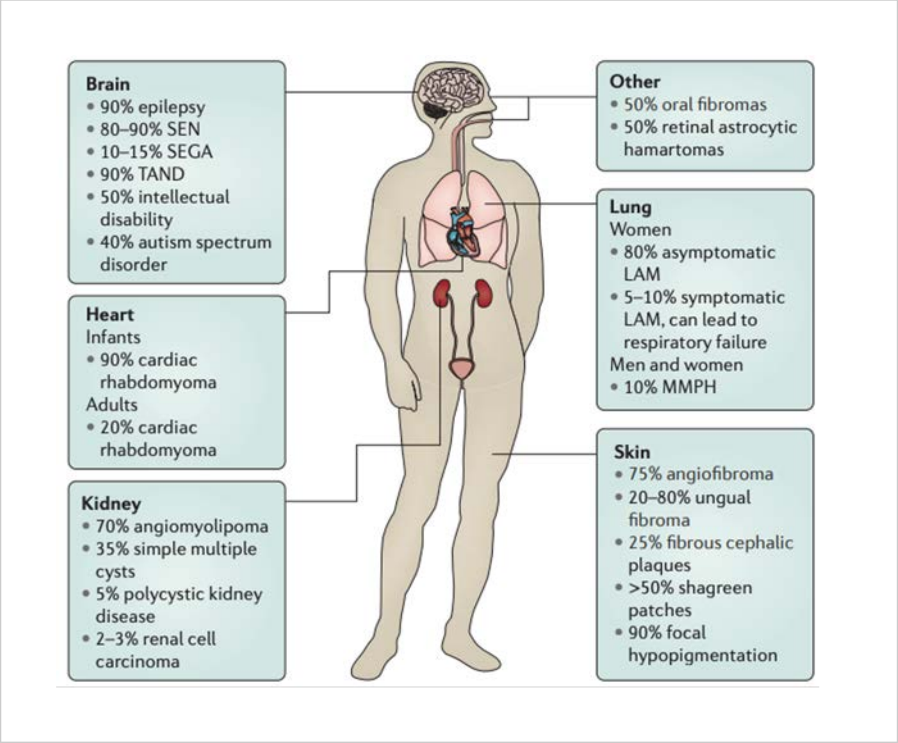
PRIORITY INDICATIONS

- Tuberous sclerosis complex
- ALS (with BMS)
- Undisclosed neurodegeneration

EXAMPLE INSITRO DISCOVERY EFFORTS

- ML-enabled analysis of multi-modal data from iPSC-derived neuronal cell models for target discovery and interrogation
- Familial mutations provide phenotypic ground truth for sporadic disease in cell models

Discovery in Tuberous Sclerosis Complex (TSC)



Brain
- 90% epilepsy
- 80–90% SEN
- 10–15% SEGA
- 90% TAND
- 50% intellectual disability
- 40% autism spectrum disorder

Heart
Infants
- 90% cardiac rhabdomyoma

Adults
- 20% cardiac rhabdomyoma

Kidney
- 70% angiomyolipoma
- 35% simple multiple cysts
- 5% polycystic kidney disease
- 2–3% renal cell carcinoma

Other
- 50% oral fibromas
- 50% retinal astrocytic hamartomas

Lung
Women
- 80% asymptomatic LAM
- 5–10% symptomatic LAM, can lead to respiratory failure

Men and women
- 10% MMPH

Skin
- 75% angiofibroma
- 20–80% ungual fibroma
- 25% fibrous cephalic plaques
- >50% shagreen patches
- 90% focal hypopigmentation

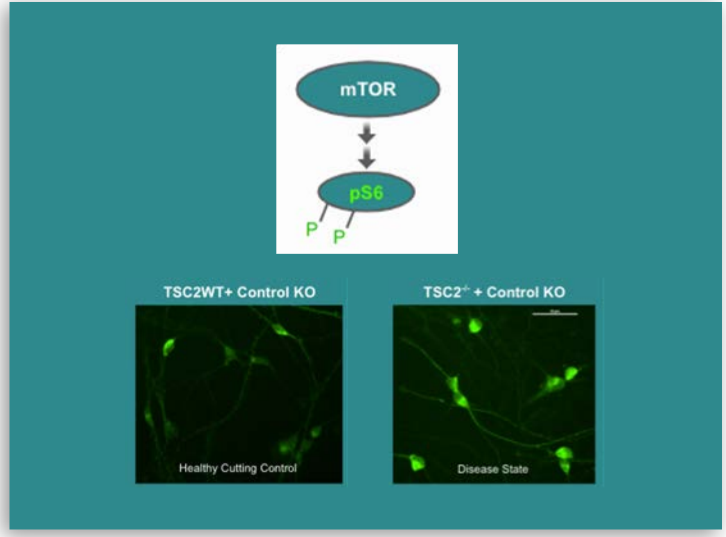
Motivation

- **Prevalence:** 50K US patients; >1M globally
- **Genetics:** *TSC1*/*TSC2* LoF leading to mTOR pathway overactivation
- **Symptoms**: Benign tumors in skin, brain and kidney leading to neuro impairment, epilepsy
- **Everolimus** approved for adjunctive use for seizures, but ineffective in ~50% of patients
- Strong advocacy

Theilmann et al. (2020) *Neuropharmacology* 180:108297; Henske et al. (2016) *Nat Rev Dis Primers*. 2:16035.

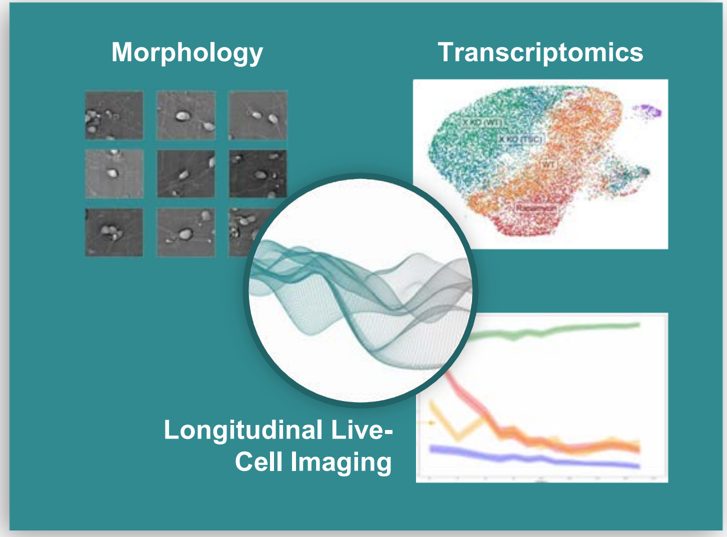
Dynamic Phenotyping for Discovery in TSC

What *others* see...



- Marker phenotypes
- Hypothesis testing

What *insitro* sees...



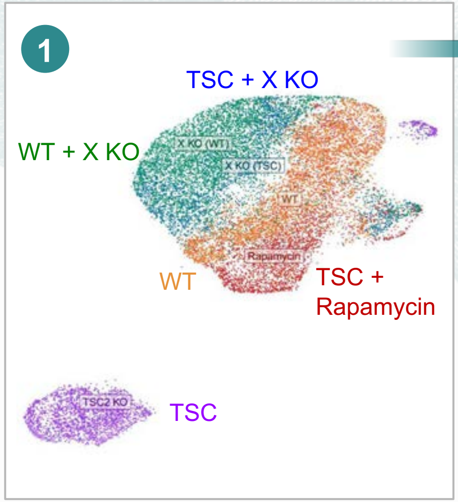
- Multi-modal and longitudinal data
- Learned phenotypes for biologic state
- Hypothesis generation, prediction

insitro

D. Koller, 2021 International TSC and LAM Research Conference, August 2021

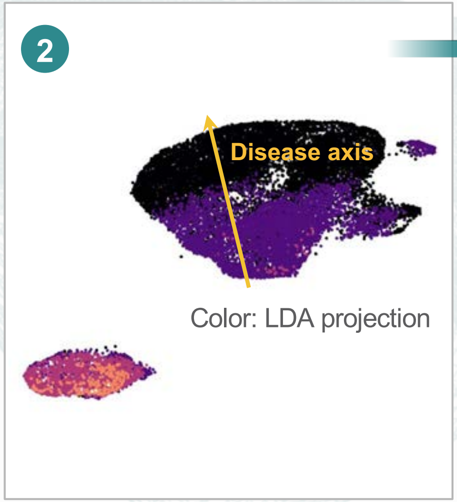
scRNAseq Model Distinguishes “Healthy” and “Sick”, and is Consistent with Patient Data

Build manifold



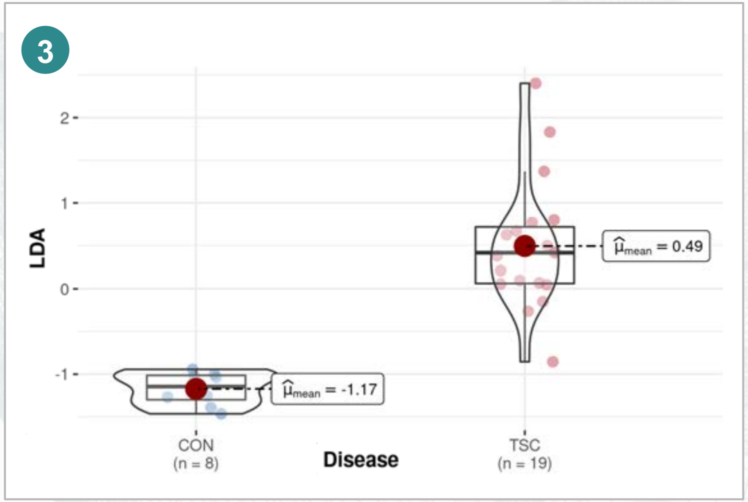
Unbiased manifold construction from scRNAseq data

Predictive model



ML (LDA) model trained to distinguish TSC$^{-/-}$ vs WT defines a disease axis

Validate in patient samples

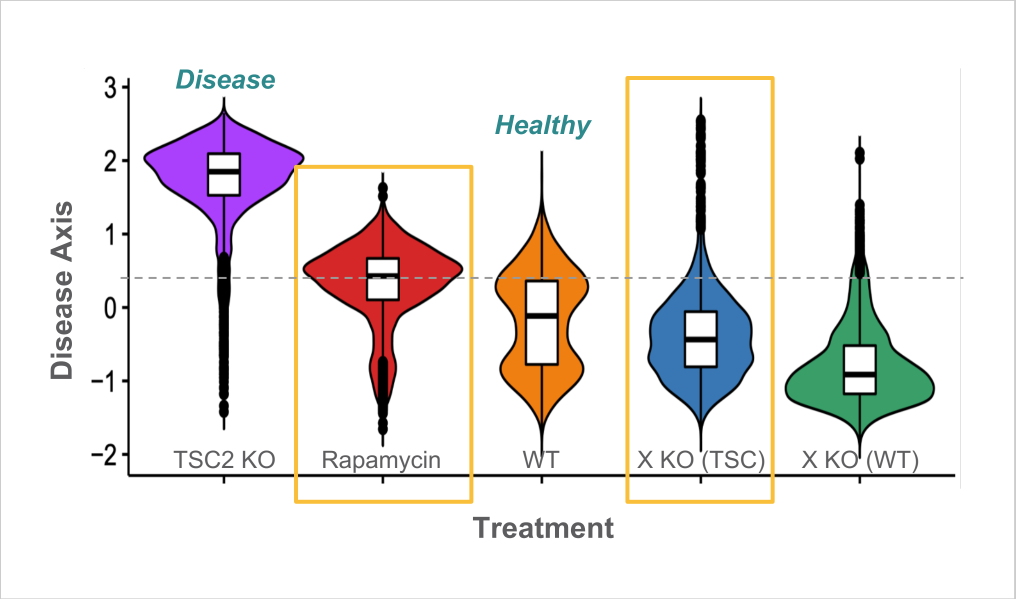


ML model applied to bulk RNAseq data from *in vivo* patient samples[1] correctly separates “disease” from “healthy”, validating disease axis

[1] (1) Bongaarts et al., 2020

scRNAseq ML Model Enables Interrogation of Chemical and Genetic Perturbations in TSC

Application of Predictive Model

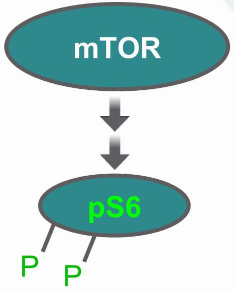


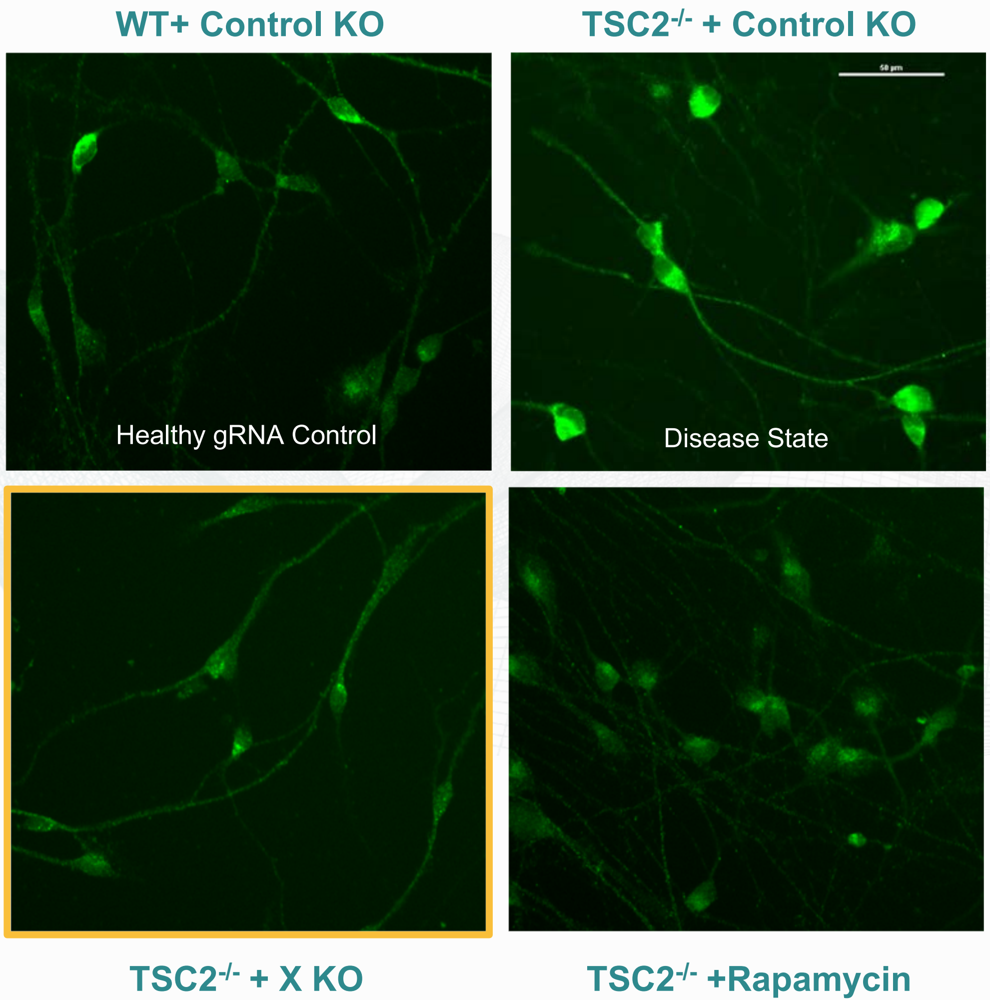
Projection of scRNAseq data onto the ML Disease Axis:

- Rapamycin shows substantial reversion, as expected
- Gene X KO also shows significant reversion, supporting its potential as an intervenable node in TSC

Tested mTOR Pathway Genes for Marker-Based Reversion

- Cells with KO of Gene X have greatly reduced pS6 signals
- Other potential targets in mTOR pathway **did not** reduce pS6

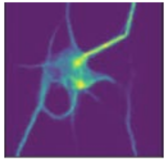




Morphology Model Predicts Reversion for Both Rapamycin and X KO Cells

Build classification model

DAPI, MAP2



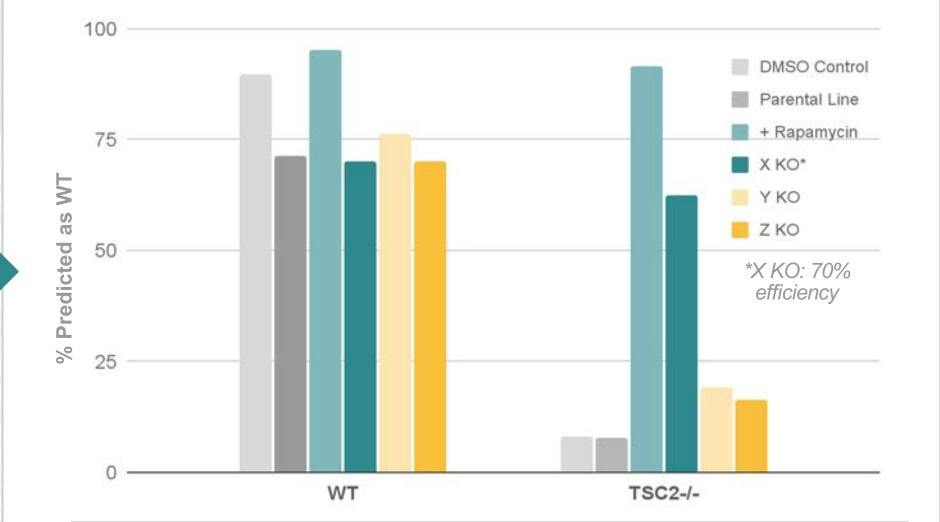
40d old neurons

TSC2$^{-/-}$ KO

WT Control

- Generate cell morphology data on TSC2$^{-/-}$ and WT with DAPI & MAP2
- Train model to classify cells as TSC2$^{-/-}$ vs WT

Use model to explore and test perturbagens

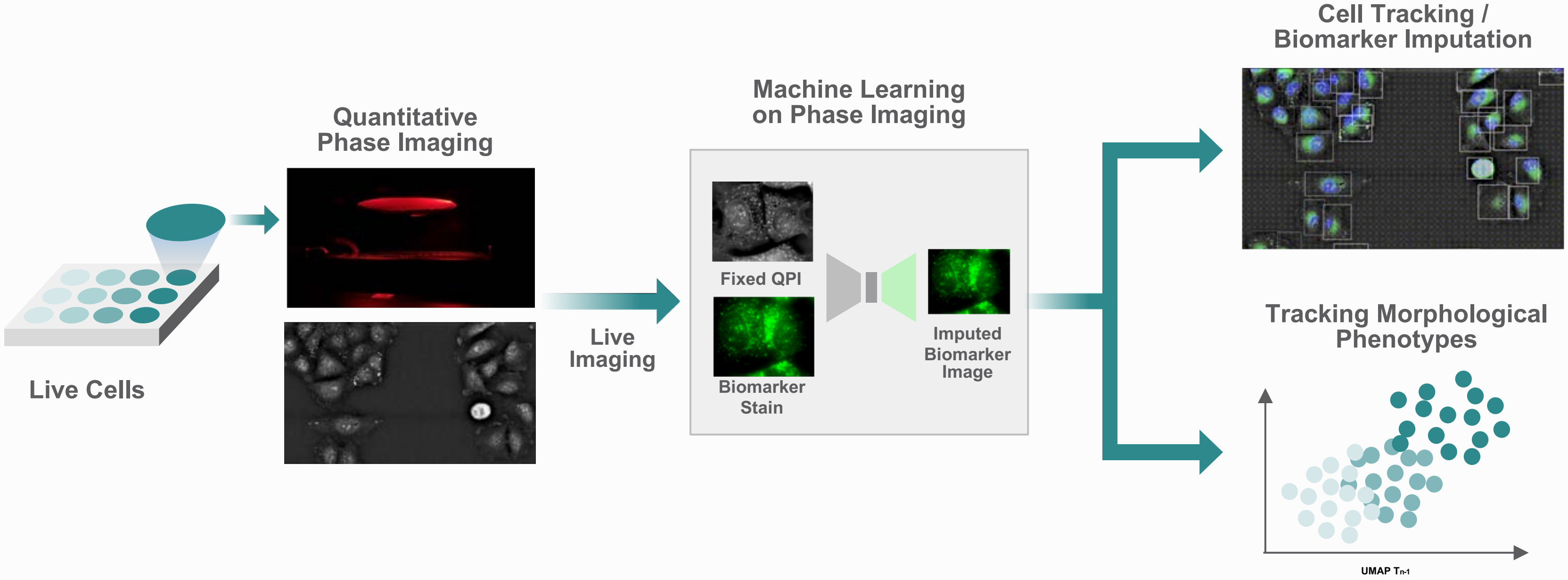


X KO: 70% efficiency

Perturbations in WT and TSC2$^{-/-}$ Cells

- TSC2$^{-/-}$ cells with **rapamycin or X KO** intervention mostly classified as “healthy”
- Interventions with KO of two other MTOR pathway genes mostly classified as “sick”, consistent with pS6 results

Novel, ML-Enabled Phenotyping Modalities Provide Insights into Cellular State and Dynamics



Live Cells

Quantitative Phase Imaging

Live Imaging

Machine Learning on Phase Imaging

Fixed QPI

Biomarker Stain

Imputed Biomarker Image

Cell Tracking / Biomarker Imputation

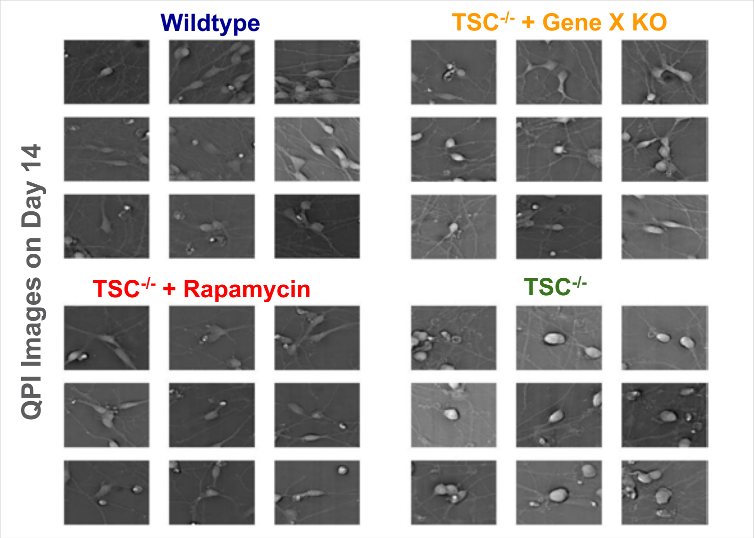
Tracking Morphological Phenotypes

Machine learning and quantitative phase imaging (QPI) enable longitudinal phenotyping and exposure response tracking in live cells

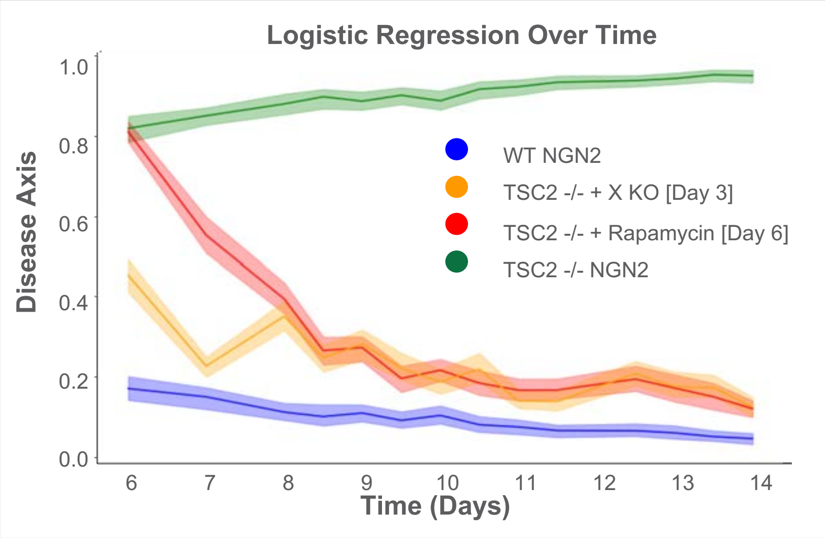
H. Marie-Nelly, Machine Learning for Computational Biology November 2021
D. Koller, British Machine Vision Conference November 2021

QPI Live Imaging Model Supports Reversion for Both Rapamycin and Gene X KO Cells

Others see: static morphology…

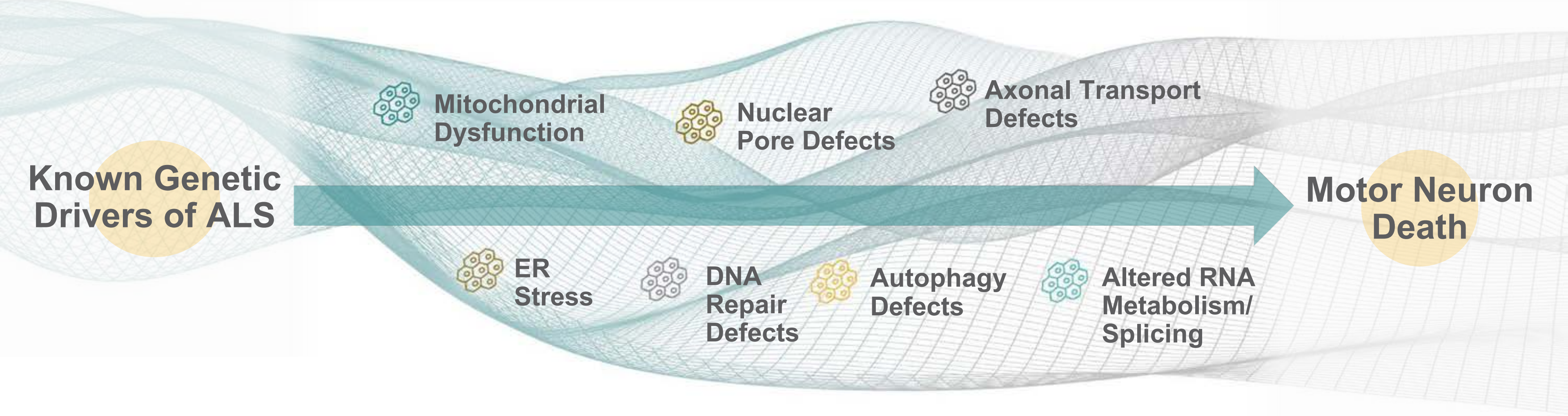


…we see live phenotypic reversion over time



ML-enabled drug discovery efforts ongoing against Gene X

Phenotypic Manifold Generation for ALS



Known Genetic Drivers of ALS

Mitochondrial Dysfunction

Nuclear Pore Defects

Axonal Transport Defects

ER Stress

DNA Repair Defects

Autophagy Defects

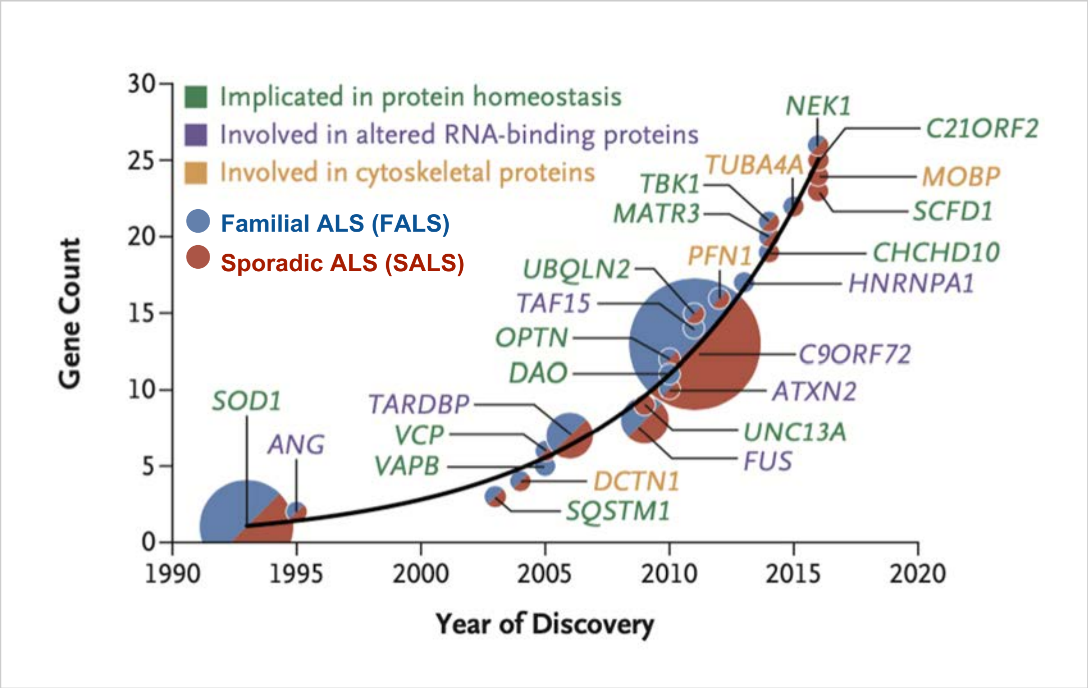
Altered RNA Metabolism/ Splicing

Motor Neuron Death

>70 engineered isogenic lines, >50 sporadic patient lines and >50 matched controls
ML-enabled automated differentiation; dense, multi-modal phenotyping;
work in collaboration with Bristol Myers Squibb

Identify conserved pathophysiology *across heterogeneous genetic causes* to define coherent patient populations and discover high-impact targets

More Than 25 Genes Have Been Reproducibly Implicated in Familial ALS, Sporadic ALS, or Both



Shared pathological and clinical features suggest a convergence of cellular and molecular events that lead to motor neuron degeneration

Source: Brown and Al-Chalabi (2017) *NEJM*. 377:162-172.

Automated iPSC Maintenance and Motor Neuron Differentiation

- **High throughput hiPSC workcells:** Hamilton Vantage Liquid handlers, Liconic rotary automated incubators and Biotek Cytation imaging platforms
- **Integrated custom methods and scheduling**: automated plate coating, feeding, passaging and imaging
- **ML-enabled characterization models:** confluency and quality (pluripotency)

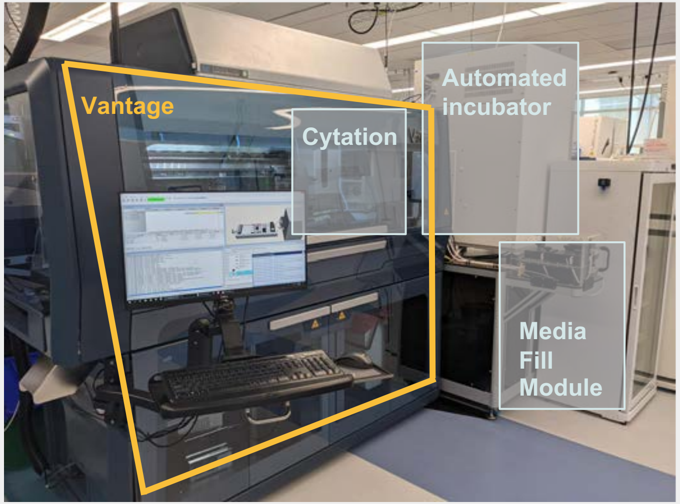


Plate Maintenance

- Custom-web based application for review of plates (multiple formats)
- Nightly whole inventory scheduled imaging
- Daily walk-away plate feed scheduling

Passaging

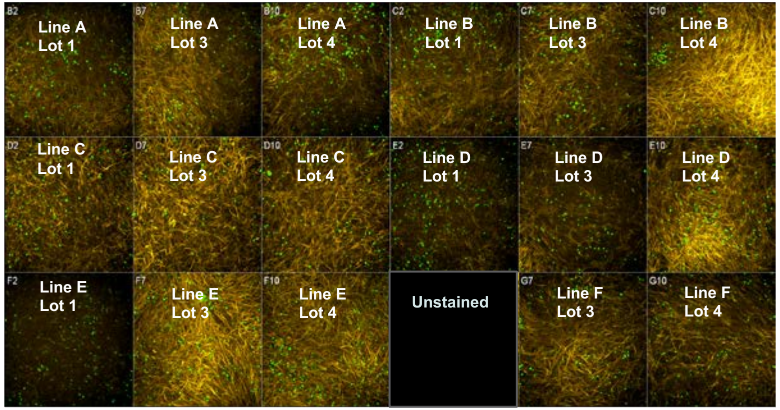
- Automated parallelized passaging of iPSC lines
- ML-informed differential seeding densities for distinct line growth rates
- Synchronization passage for differentiation

Characterization

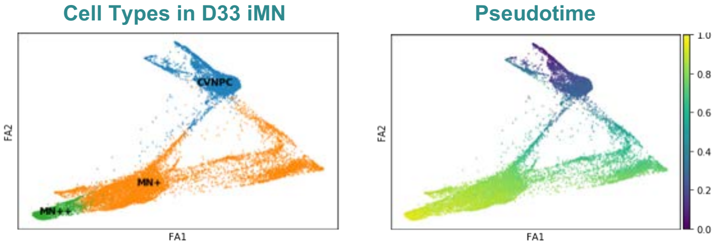
- Automated syncing to Amazon s3
- Automated ML-enabled characterization of confluency, growth rates and iPSC quality (pluripotency)
- Image-based cell counting method for cell number normalization

AutoMN Differentiation is Robust and Reproducible Across iPSC Lines

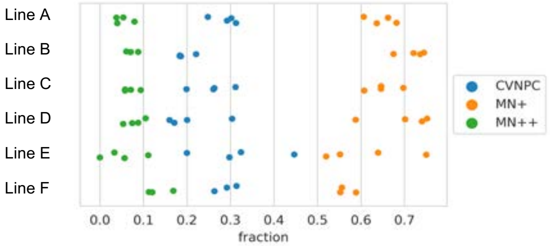
6 iPSC Lines Across 4 Replicate Runs



sc-RNA seq analysis

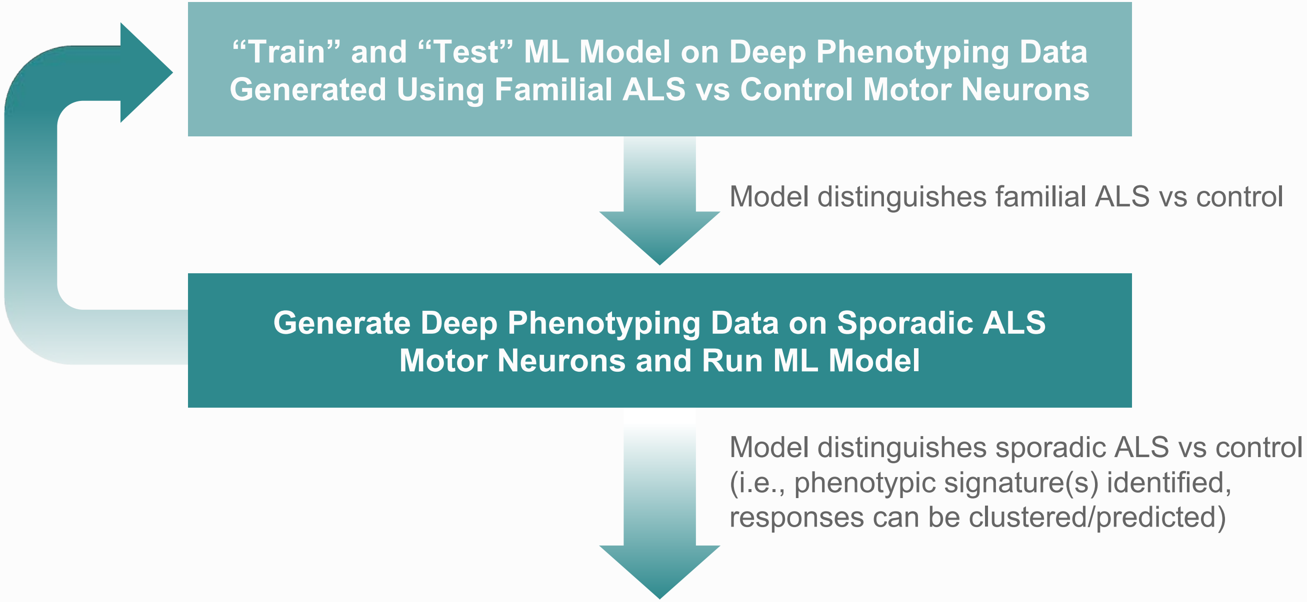

Cell Types in D33 iMN

Pseudotime



- Consistent generation of iMNs across 24 unique differentiations with rich QC time course data
- Pseudotime analysis of differentiation trajectories from each lot at Day 33 reflect high confidence in iMN clustering
- sc-RNA seq clustering of iMNs at Day 33 shows highly reproducible iMN generation across diffs and lines
- All 6 control lines closely resemble CDI motor neurons (whole transcriptome analysis)

Development of a Predictive Cellular Model of ALS



“Train” and “Test” ML Model on Deep Phenotyping Data Generated Using Familial ALS vs Control Motor Neurons

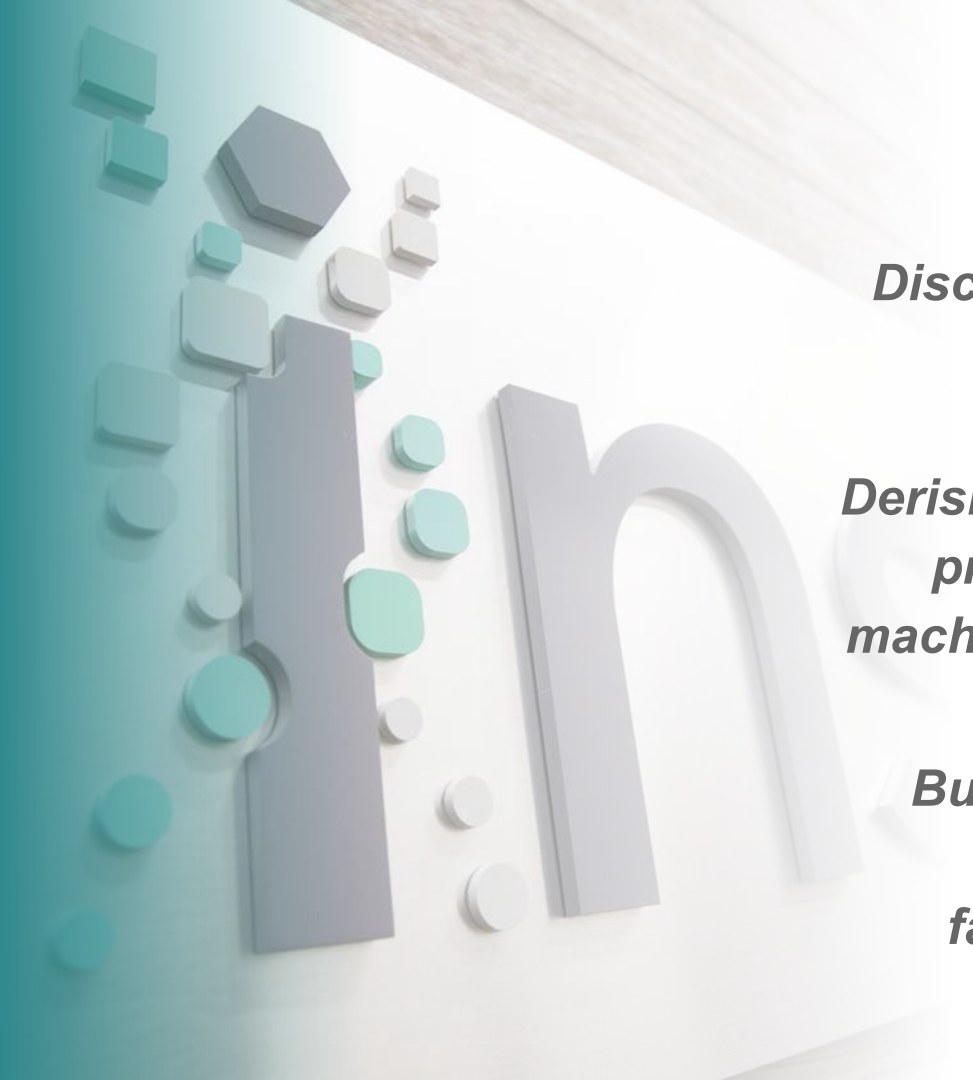
Model distinguishes familial ALS vs control

Iterative

Generate Deep Phenotyping Data on Sporadic ALS Motor Neurons and Run ML Model

Model distinguishes sporadic ALS vs control (i.e., phenotypic signature(s) identified, responses can be clustered/predicted)

Proceed to
Target Identification
(Genetic and/or chemical screens)



insitro’s Aspiration

Discover and develop transformative medicines for patients

Derisk and accelerate R&D through predictive models based on machine learning and data at scale

Build a data-enabled flywheel that allows us to do this faster and better over time